Large Deviation Principle for Markov Chains in Discrete Time

Arnaud de la Fortelle, Guy Fayolle

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large Deviation Principle for Markov Chains in Discrete Time

Arnaud de La Fortelle et Guy Fayolle

N° 3791

Novembre 1999

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3791--FR+ENG

Large Deviation Principle for Markov Chains in Discrete Time

Arnaud de La Fortelle* et Guy Fayolle†

Thème 1 — Réseaux et systèmes
Projet MEVAL

Rapport de recherche n° 3791 — Novembre 1999 — 34 pages

Abstract: Let E be a denumerable state space, X be an homogeneous Markov chain on E with kernel P. Then the chain X verifies a *weak* Sanov's theorem, i.e. a weak large deviation principle holds for the law of the pair empirical measure. In our opinion this is an improvement with respect to the existing literature, insofar as the LDP in the Markov case often requires either the finiteness of E, or strong uniformity conditions, which important classes of chains do not verify (e.g. classical queueing networks with bounded jumps). Moreover this LDP holds for *any* discrete state space Markov chain, possibly non ergodic.
The result is obtained by a new method, allowing to extend the LDP from a finite state space setting to a denumerable one, somehow like a the projective limit approach. The analysis presented here offers some by-products, among which an analogue of Varadhan's integral lemma and, under restrictive conditions, a contraction principle leading directly to a weak Sanov's theorem for the one–dimensional empirical measure.

Key-words: Large deviations, Markov chain, pair empirical measure, Sanov, entropy, information, cycle.

(Résumé : tsvp)

The work of the first author is a part of a PhD thesis supported by the École Nationale des Ponts et Chaussées and by the French Ministère de l'Équipement.

* E-mail address: Arnaud.De_La_Fortelle@inria.fr
† E-mail address: Guy.Fayolle@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : 01 39 63 55 11 - International : +33 1 39 63 55 11
Télécopie : (33) 01 39 63 53 30 - International : +33 1 39 63 53 30

Grandes déviations pour les chaînes de Markov en temps discret

Résumé : Soit X une chaîne de Markov homogène sur un espace E dénombrable. Alors un théorème de Sanov *faible* s'applique à X, i.e. la mesure empirique d'ordre 2 vérifie un principe faible de grandes déviations (PGD). Cette proposition est une amélioration des résultats actuels. En effet, en général, les PDG soit supposent la finitude de E, soit imposent sur X une forte condition d'uniformité, qui exclut d'importantes classes de chaînes et notamment les réseaux de files d'attente à sauts bornés. Par ailleurs ce PGD est valide pour *toute* chaîne de Markov à espace d'états discret, sous la seule hypothèse d'irréductibilité.
Ce résultat est le fruit d'une nouvelle approche, qui permet d'étendre des PGD en passant d'espaces d'états finis à un espace d'états dénombrable. Il faut noter qu'ici une suite de PGD forts (ils le sont nécessairement sur un espace d'états fini par exemple) implique seulement un PGD faible, contrairement à ce qui se passe habituellement pour les limites projectives.
On obtient aussi un certain nombre de corollaires, parmi lesquels un analogue du lemme intégral de Varadhan ou encore, sous des conditions assez restrictives, un principe de contraction qui entraine immédiatement le théorème de Sanov faible pour la mesure empirique d'ordre 1. Soulignons enfin que la méthode s'étend aisément à des contextes voisins : chaines réductibles, temps continu, espaces d'états plus généraux.

Mots-clé : Grandes déviations, chaîne de Markov, mesure empirique, Sanov, entropie, information, cycle.

1 Introduction and main ideas

Large deviations theory is a powerful tool to analyze the behaviour of stochastic processes in the long run, off the Central Limit Theorem regime. Estimates have been used in number of applications, some general results being specialized in various ways to fit each case.

The goal of our study is to propose a proof of a LDP for Markov chains defined on a countable space. This general framework includes most of the classical queueing networks with bounded jumps and Markovian evolution.

Assumptions, techniques and bounds When the state space is finite, results are well known and the rate function is expressed as an *entropy function*, the so-called *Kullback–Leibler information* [27]. In fact, the link between large deviations and entropy was frequently used [22, 4, 7, 33], and earlier works on information theory already emphasized the equivalence of entropy to a *mean information gain* [37, 24]. This interpretation often proved very useful [31, 30, 25, 34, 5] and is briefly recalled in Appendix B.1.
Since its infancy [19], the standard theory (e.g. [16, 14, 15]), interested in Markov chains defined on Polish spaces, does impose a pretty strong uniformity condition of the following type:

There exist integers $M \geq 1$ and $N, 0 < l \leq N$ such that, for all states x, y,

$$P^{(l)}(y,.) \leq \frac{M}{N} \sum_{m=1}^{N} P^{(m)}(x,.)\,, \tag{1.1}$$

where $P^{(m)}(x,.)$ denotes the m–step transition probability, given the initial state x. Alas, this condition is very restrictive. As a rule, the M/M/1 queue, Jackson networks and most of the discrete event systems with downward-bounded jumps *do not satisfy it.* Although refinements were proposed [23, 11, 41, 32], they still do not suffice for our purpose. General bounds have been obtained by considering kernels of the form

$$K_\xi(x,A) = \int_A e^{\langle \xi, f(y) \rangle} P(x,\, \mathrm{d}y),$$

for all bounded continuous functions f and all continuous linear functionals ξ (see [10, 11, 17, 12]). In this respect, it is worth quoting the study in [12], which nicely brings out the fact that the sole irreducibility is necessary for the lower bound to hold (even for an unbounded functional f).

In [18], the authors consider the quantity $\mathbb{P}\left[L_n \in B(\mu,\varepsilon), X_n \in C\right]$, where $B(\mu,\varepsilon)$ and C denote respectively a ball and a *small* set. Then, by means of regeneration arguments, it is shown that the sets $B(\mu,\varepsilon)$ are subject to an exponential decay, whence general lower and upper bounds follow; but the problem of getting rid of the set C still remains.
At that moment, it is important to point out the existence of strong technical subtleties, which concern the coincidence of the rate functions for the upper and lower bounds. This point is illustrated in Section 3.5.

In network models, the families of probability measures of interest are rarely exponentially tight, so that, in many applications, a strong LDP does not hold. Therefore, we chose to concentrate on the weak LDP. Our approach (improved since the first release [13]) allows to get the upper bound with the *same* rate function for both compact and open sets. Moreover proofs are short, constructive, and they extend to special situations, e.g. reducible chains, continuous time, etc.

Our goal is to analyze the following object (see [21, 22, 23]).

Definition 1.1 (pair empirical measure) *The pair empirical measure is defined by*

$$L_n(\omega) \stackrel{\text{def}}{=} \frac{1}{n}\left(\sum_{i=1}^{n-1}\delta_{X_i(\omega),X_{i+1}(\omega)} + \delta_{X_n(\omega),X_1(\omega)}\right) \in M_s(E^2),$$

where $M_s(E^2)$ is the set of balanced *measures* [1]*. The last transition*[2] *(X_nX_1) will be referred to as a "ghost transition".*

The organization of the paper is streamlined. Starting from the finite state space LDP (Section 2.1), an extension to *locally Markov processes* is obtained (Section 2.2), which, by using continuity properties of the rate function (Appendix B), leads to the weak Sanov's theorem (Sections 2.3–2.4).

Theorem 1.1 (Generalized Sanov's theorem) *Let X be an irreducible Markov chain with kernel P. Then the pair empirical measure L_n satisfies a weak LDP with rate function $H(.\|P)$, i.e. for all open sets $O \subset M_s(E^2)$ and all compact sets*

[1] The notation used throughout the paper is given in Appendix D.

[2] This transition is added for L_n to stay in $M_s(E^2)$ which is the "natural" set-up for the LDP. When Sanov's theorem is derived from the LDP for the single empirical measure, the ghost transition is chosen to be (X_0X_1) with X_0 arbitrary. Then $L_n \in M_1(E^2)$, but the functional is the same and is infinite outside $M_s(E^2)$).

$K \subset M_s(E^2)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq - \inf_{A\in O} H(A\|P), \tag{1.2}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A\in K} H(A\|P), \tag{1.3}$$

where the entropy function H (see Appendix B) is defined by

$$H(A\|P) \stackrel{\text{def}}{=} \sum_{i,j\in E} a_{ij} \log\left(\frac{a_{ij}}{a_i P_{ij}}\right).$$

This LDP holds without any restriction on the initial distribution. In particular, no uniformity condition is imposed.

For the upper bound equation (1.3), the analysis of the entropy function H (i.e. the *rate function*) in Section B shows important differences, that change the nature of the LDP. Indeed, in the countable state space case, H is lower semi-continuous and the level sets

$$\Psi(\alpha) \stackrel{\text{def}}{=} \left\{A \in M_s(E^2) : H(A\|P) \leq \alpha\right\}$$

are *not necessarily compact*. This means at once a *full* LDP cannot be achieved, but we prove nevertheless that a *weak* LDP holds.

The main idea to get the lower bound equation (1.2) (see Section 2.4) is to construct successive finite approximations of a measure A, remarking that the lower bound is easily proved for finite support measures. Nonetheless, convergence problems might occur, since H is not continuous. To circumvent them, we analyze in Appendix C the link between balanced measures and cycles, and the continuity of the entropy is discussed in Appendix B.

2 Sanov's theorem

2.1 The finite case

Large deviation principles in a finite state space have been proved quite a while ago, since famous Sanov's theorem [39] was first published in 1957, in Russian. Convexity and eigenvalues properties are the major tools to get the result [35], but combinatorial estimates are sufficient [8]. In our opinion, good books presenting methods and major references are [16, 15]. The rate function H, sometimes called the *Kullback–Leibler divergence*, became a classical quantity, since its appearance in information theory in 1951 [27, 24],

Theorem 2.1 (Basic Sanov's theorem) *Let X be a Markov chain on a finite state space E. The pair empirical measure L_n satisfies a LDP with the good rate function $H(.\|P)$.*

Proof : Only the main lines will be presented.

Let $F \subset M_s(E^2)$ be closed. The upper bound relies on the following combinatorial inequalities (see [13])

$$\mathbb{P}\left[L_n = A\right] \leq e^{\varepsilon_n} \prod_{i,j \in E} \frac{(na_i)!}{(na_{ij})!} P_{ij}^{a_{ij}} \leq e^{-nH(A\|P)+\varepsilon_n}, \tag{2.1}$$

where $\varepsilon_n = o(n)$ and is independant of A. Since there are at most $(n+1)^{|E|^2}$ empirical measures L_n of length n, we have

$$\log \mathbb{P}\left[L_n \in F\right] \leq |E|^2 \log(n+1) - n \inf_{A \in K} H(A\|P) + \varepsilon_n,$$

and hence the result.

Let $O \subset M_s(E^2)$ be open, $A \in O$ with $H(A\|P) < \infty$, so that $A \in M_s(G)$, where G is the graph associated to P, remarking that these assumptions are sufficient, as far as the infimum of the lower bound is concerned.
One will use a classical exponential change of measure, via the martingale

$$M_n \stackrel{\text{def}}{=} \exp \sum_{i=1}^{n-1} h(X_i, X_{i+1}), \quad \text{where } h(i,j) = \log \frac{A_{ij}}{P_{ij}}. \tag{2.2}$$

So, M_n defines a new probability $\mathbb{P}^*$, under which X is a Markov chain with transition matrix A (the function h is finite because $A \in M_s(G)$). Denoting $\mathrm{Supp}(A)$ by G', we have, for any open neighborhood $\mathcal{V}_A \subset O$ of A,

$$\begin{aligned}
\frac{1}{n} \log \mathbb{P}\left[L_n \in \mathcal{V}_A\right] &= \frac{1}{n} \log \mathbb{E}^*\left[\mathbb{1}_{\{L_n \in \mathcal{V}_A\}} M_n^{-1}\right] \\
&\geq -\sup_{A' \in \mathcal{V}_A \cap M_s(G')} \sum_{(i,j) \in G'} a'_{ij} h(i,j) && (2.3) \\
&\quad + \frac{1}{n} \log \mathbb{E}^*\left[\mathbb{1}_{\{L_n \in \mathcal{V}_A\}} e^{-h(X_n, X_1)}\right]. && (2.4)
\end{aligned}$$

The term in (2.4) tends to 0, since X_n is ergodic, (so that L_n converges to A in probability), and h is bounded on $M_s(G')$. When $\mathcal{V}_A$ is sufficiently close to A, the

term in (2.3) does not differ substantially from $H(A\|P)$, due to the continuity of the mapping
$$A \to \sum_{(i,j)\in G'} a_{ij} h(i,j).$$
Since $\mathbb{P}[L_n \in O] \geq \mathbb{P}[L_n \in \mathcal{V}_A]$ for all A, the lower bound is proved. ■

Actually A might be not irreducible, in which case $\mathbb{P}^*$ does not have a unique stationary measure. Thus one must first prove the lower bound for any irreducible A, and then extend the result to any A, by a continuity argument, using Proposition C.1 and Proposition B.2.

M_n is only defined on Supp(A), which is absorbing under $\mathbb{P}^*$. This would be enough to conclude, provided that the initial distribution ν satisfies. $\nu\big(\mathrm{Supp}(A)\big) > 0$, but this condition is not true in general. Therefore, one chooses h to be null outside Supp(A), so that M_n is defined everywhere.
Under $\mathbb{P}^*$, X behaves like P outside Supp(A). From the irreducibility of P, there exists n_0 with $\mathbb{P}^*[X_{n_0} \in \mathrm{Supp}(A)] > 0$. It suffices now to consider the chain for $n \geq n_0$, so that the bounds still hold, as the finite part $(X_n, n \leq n_0)$ does not play any role.
In fact, the way of defining X under $\mathbb{P}^*$ outside the support of A is not crucial, so that the next section is in some sense quite natural.

2.2 Extension to locally Markov processes

Here one proposes to extend Theorem 2.1 to a countable state space. This leads first to make a slight generalization of Theorem 2.1 to the case of *substochastic* transition matrices.

Definition 2.1 (Locally Markov process) *A process X_n is said to be locally Markovian (and homogeneous) on $E' \subset E$, with transition matrix P, if*
$$\mathbb{P}[X_{n+1} = y | \mathcal{F}_n \cap \{X_n = x\}] = P_{xy}, \quad \forall x, y \in E',\ \forall n \geq 0, \tag{2.5}$$
where $\mathcal{F}_n$ is the natural filtration of X_n. For the sake of consistency, one requires the probability of ever reaching E' to be strictly positive.

Note that there is no restriction on the initial probability distribution, so that $\mathbb{P}[X_n \in E']$ may behave arbitrarily.[3]

[3] For instance, if X is a random walk on $\mathbb{Z}$ and E' is finite, a heavy tailed initial probability distribution disturbs a LDP for $\mathbb{P}[X_n \in E']$. Fortunately this is not the case for the empirical measure.

Using the combinatorics of (2.1) for the upper bound, taking the value 1 the unknown probabilities, and the above change of measure for the lower bound, one easily obtains the following extension of Theorem 2.1.

Theorem 2.2 *Let X be a locally Markovian process on a finite set E'. The pair empirical measure L_n satisfies the LDP lower bound on $M_s(E'^2)$ and, restricted to E', satisfies the LDP upper bound, both with the good rate function $H_{E'}(.\|P)$. More formally, for all open $O \subset M_s(E'^2)$ and all closed $F \subset t_{E'}(M_s(E^2))$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq - \inf_{A\in O} H_{E'}(A\|P), \tag{2.6}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[t_{E'}(L_n) \in F\right] \leq - \inf_{A\in F} H_{E'}(A\|P), \tag{2.7}$$

where $t_{E'}(A) \stackrel{\text{def}}{=} (a_{ij})_{i,j\in E'}$ is the truncation[4] *on E'.*

Although Theorem 2.2 is not a LDP, it looks very much like a kind of *projective limit*. Nonetheless, the projection of a Markov chain onto some subset is no longer a Markov chain, and consequently the projective limit method should undergo ad hoc modifications. Our approach is qualitatively different, since a *full* LDP for finite state spaces becomes in general a *weak* LDP for the limiting countable state space.

The set $t_{E'}(M_s(E^2))$ is *not equal* to $M_s(E'^2)$, nor to the sets of all measures on E'^2 with mass less than 1. The above formulation is quite natural, and avoids tedious problems related to the existence of paths entering or leaving each state a fixed number of times.

2.3 Upper bound

The heart of the extension consists of three properties. First the compacity of the set of measures, second the fact that a Markov chain is locally Markovian on any finite set, and third the lower semi-continuity of the rate function H.

Theorem 2.3 *Let X be a Markov chain on a countable state space E, with transition matrix P. The pair empirical measure L_n satisfies the LDP upper bound on compact sets with the rate function $H(.\|P)$, that is, for all compact sets $K \subset M_s(E^2)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A\in K} H(A\|P). \tag{2.8}$$

[4] For the sake of simplicity, we shall denote by $t_{E'}$ both the transition matrices and the restriction of measures on E' or on E'^2.

Proof : Let $x < \inf_K H(.\|P)$ be fixed. Then there exists a finite subset $E' \subset E$ such that

$$\inf_{A \in K} H_{E'}(A\|P) \geq x. \tag{2.9}$$

Suppose this is not true. Then, for all finite E_n, there exists $A(n) \in K$ such that $H_{E_n}(A(n)\|P) < x$. Choosing $E_n = \{1, \ldots, n\}$, this writes

$$\liminf_{n \to \infty} H_{E_n}(A(n)\|P) \leq x.$$

Since K is compact, there exists a subsequence $A(n)$ converging to A in K. Of course $t_{E_n}(A(n))$ also converges to A, because E_n increases to E. Applying Definition B.2, we have

$$H_{E_n}(A(n)\|P) = H(t_{E_n}(A(n))\|P).$$

By Lemma B.1, H is lower semi-continuous, so that

$$H(A\|P) \leq \liminf_{n \to \infty} H(t_{E_n}(A(n))\|P) \leq x.$$

But this contradicts $x < \inf_K H(.\|P)$. Hence, (2.9) is true for some finite subset $E' \subset E$. As X is obviously locally Markovian on E', Theorem 2.2 yields

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[t_{E'}(L_n) \in t_{E'}(K)\right] \leq - \inf_{A \in t_{E'}(K)} H_{E'}(A\|P).$$

Now

$$\inf_{A \in t_{E'}(K)} H_{E'}(A\|P) = \inf_{A \in K} H_{E'}(t_{E'}(A)\|P) = \inf_{A \in K} H_{E'}(A\|P) \leq -x.$$

Since $L_n \in K$ implies that $t_{E'}(L_n) \in t_{E'}(K)$, we obtain

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq -x,$$

which this is true for any $x < \inf_K H(.\|P)$. Theorem 2.3 is proved. ■

2.4 Lower bound

The extension for the lower bound is simple, by using continuity properties of H.

Theorem 2.4 *Let X be a Markov chain on a countable state space E, with transition matrix P. The pair empirical measure L_n satisfies the LDP lower bound with the rate function $H(.\|P)$, that is, for all open sets $O \subset M_s(E^2)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq - \inf_{A\in O} H(A\|P). \tag{2.10}$$

Proof : Let E' be a finite subset of E and $A \in M_s({E'}^2) \cap O$. Since O is open, there exists an open neighborhood $\mathcal{V}_A \subset M_s({E'}^2)$ of A included in O. Applying Theorem 2.2 for such a measure, one has

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in \mathcal{V}_A\right] \geq -H_{E'}(A\|P).$$

But $H_{E'}(A\|P) = H(A\|P)$ and $\mathcal{V}_A \subset O$, so that the lower bound is satisfied for any measure $A \in O$ having a finite support. This reads, for such A,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq -H(A\|P). \tag{2.11}$$

Let us fix now $A \in O$. By Proposition B.3, there exist measures $A(m)$ with finite support, converging to A, and ensuring also the convergence of $H(A(m)\|P)$ to $H(A\|P)$. Applying (2.11), with A replaced by $A(m)$, results in

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq -H(A(m)\|P) \overset{m\to\infty}{\longrightarrow} -H(A\|P).$$

Thus (2.11) is valid for any measure $A \in O$ and the proof of Theorem 2.4 is concluded. ∎

> **Remark 1 :** The change of measure works in fact even for an infinite state space. In this case, problems encountered have a twofold aspect. First, the method works only for irreducible A: no matter which approach is taken, the study of continuity can never be avoided, if the goal is to extend the lower bound to all $A \in O$. Secondly, when $\mathrm{Supp}(A)$ is countable, the supremum in equation (2.3) may be infinite (because h itself can be unbounded), and therefore it is necessary to place some restrictions on A.

3 Links and transformations of weak LDP

3.1 Another derivation via Ruelle–Lanford functions

In a beautiful paper [29], J. Lewis and C. Pfister analyze, in a very broad setting, large deviation phenomena from a thermodynamical point of view. It is interesting to note that their *vague* LDP corresponds exactly to our[5] *weak* LDP.

In this respect, a pragmatic question arises: could the weak LDP be obtained in a more direct way by means of Ruelle–Lanford functions? More precisely, assume one shows $H(A\|P)$ is a Ruelle–Lanford function: then Theorem 3.1 of [29] yields immediately the weak LDP. This is true, but (in agreement with the law saying that miracles are rare events!), the proof is in essence not simpler as the one given in Sections 2.3 and 2.4. Indeed Theorem 1.1 can be derived without the notion of local Markov processes, but at the expense of a deeper analysis of the change of measure. For the sake of completeness, we prove the following result.

Theorem 3.1 *The entropy is a Ruelle-Lanford function for the empirical measure* L_n.

Proof : The notation is as in Theorem 2.1, but here E is infinite. When $A \in M_s(G)$ is an irreducible measure with a finite support, the line of argument which leads to the lower bound still holds, but (2.3) also yields the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in \mathcal{V}_A\right] \leq - \inf_{A' \in \mathcal{V}_A \cap M_s(G')} \sum_{(i,j)\in G'} a'_{ij} h(i,j). \tag{3.1}$$

By continuity on $M_s(G')$ of the mapping $A \to \sum_{i,j} a_{ij} h(i,j)$, for all $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}_A$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in \mathcal{V}_A\right] \leq -H(A\|P) + \varepsilon. \tag{3.2}$$

Let now A be arbitrary, $\mathcal{V}_A$ a neighborhood of A, and $K \subset \mathcal{V}_A$. In Propositions C.1 and C.2, one shows the set of irreducible measures with finite support to be dense in $M_s(G)$. Consider all such $A' \in \mathcal{V}_A$ and their corresponding neighborhoods $\mathcal{V}_{A'}$, subject to equation (3.2). Then $\mathcal{V}_A$ and K can be covered by a collection of

[5] The terminology [29] follows O'Brien and Vervaat [36], while we use Deuschel and Strook [16] vocabulary.

sets $\mathcal{V}_{A'}$, from which, since K is compact, we can extract a finite covering such that $K \subset \bigcup_{A' \in C(\varepsilon)} \mathcal{V}_{A'}$. Hence, by equation (3.2),

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A' \in C(\varepsilon)} H(A' \| P) + \varepsilon. \tag{3.3}$$

Therefore, by the lower semi-continuity of H and equation (3.3),

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq -H(A \| P) + 2\varepsilon, \tag{3.4}$$

and the local upper bound is proved.

Using the lower semi-continuity of H, Proposition B.3 and the density of irreducible measures with finite support, on can prove in a similar way a local lower bound for any measure A, exactly the reverse inequality of (3.2). Finally, $H(A\|P)$ is a Ruelle–Lanford function for $\mathbb{P}\left[L_n \in .\right]$ in the Hausdorff topological space $M_s(E^2)$, so that, by Theorem 3.1 in [29], the weak Sanov's theorem is proved. ■

Comparison As asserted at the beginning of this subsection, the two proofs have roughly the same complexity.

- In Section 2, one needs a detailed analysis of finite substochastic transition matrices, which form the locally Markov processes, and the result is a corollary of the finite state space LDP.
- The method proposed here is an alternative to that of Section 2. Since $M_s(E^2)$ might be not locally compact, obtaining the upper bound is not really simpler than the LDP upper bound. Also, though further properties are explicitly used (e.g. continuity, density irreducible measures with finite support), no more work is needed because they are indispensable for the lower bound.

The choice of the method should be made according to the respective complexity of the bounds (3.2) and (2.7).

3.2 Varadhan's integral lemma and contraction principle

The next corollary[6], which is the analog of Varadhan's integral lemma, is given in [29].

[6]This could also be deduced easily from the change of measure. In fact the proof already relies on the fact that M_n^{-1} is a continuous functional of L_n.

Theorem 3.2 *Let $g : M_s(E^2) \to \mathbb{R}$ be a continuous function. Then, for all open sets O and compact sets K,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathbb{1}_{\{L_n\in O\}} e^{ng(L_n)}\right] \geq - \inf_{A\in O} \{H(A\|P) - g(A)\}, \tag{3.5}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathbb{1}_{\{L_n\in K\}} e^{ng(L_n)}\right] \leq - \inf_{A\in K} \{H(A\|P) - g(A)\}. \tag{3.6}$$

Hereafter, we prove a somewhat restricted contraction principle, which contains interesting by-products (see Theorem 3.4).

Theorem 3.3 (contraction) *Let $f : M_s(E^2) \to F$ be a continuous functional such that $f^{-1}(K)$ is compact, for any compact set $K \subset F$. Then $\mathbb{P}[f(L_n) \in .]$ satisfies a weak LDP with the rate function J.*

$$J(y) \stackrel{\text{def}}{=} \inf_{A\in f^{-1}(y)} H(A\|P) \tag{3.7}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}[f(L_n) \in O] \geq - \inf_{\mu\in O} J(y), \tag{3.8}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[f(L_n) \in K] \leq - \inf_{\mu\in K} J(y). \tag{3.9}$$

Proof : The positivity of J, like equations (3.8)–(3.9), are immediate. The only difficulty concerns the lower semi–continuity of J.
Letting $\mu(n) \stackrel{n\to\infty}{\longrightarrow} \mu$, the set

$$K(n_0) = \{\mu(n)\}_{n\geq n_0} \cup \{\mu\}$$

is compact, and so is $f^{-1}\big(K(n_0)\big)$. By a finite covering argument and the lower semi–continuity of H, is it not difficult to prove the inequalities

$$\begin{aligned} \liminf_{n\to\infty} J(\mu(n)) &= \liminf_{n\to\infty} \inf_{A\in f^{-1}(\mu(n))} H(A\|P) \\ &\geq \inf_{A\in f^{-1}(\mu)} H(A\|P) = J(\mu). \end{aligned}$$

■

Instantiating for example $f(A) = A^{(1)}$, one can checks in this case that all assumptions of Theorem 3.3 are satisfied: hence, the contraction is valid and there is a LDP for the one–dimensional empirical measure.

Theorem 3.4 *Let X be an irreducible Markov chain with kernel P. Then the empirical measure satisfies a weak LDP with the rate function*

$$I_0(\mu\|P) \stackrel{\text{def}}{=} \inf_{A^{(1)}=\mu} H(A\|P). \tag{3.10}$$

The identification of I_0 in equation (3.10) does not require any particular condition (see Appendix 3.5), so that the weak Sanov's theorem in the one–dimensional case is really an easy consequence of Theorem 1.1.

3.3 Links with a full LDP

There is no way of improving our results toward a full LDP, as shown in the remark in Section B.1. We emit the conjecture that the problem of existence of a full LDP reduces to the analysis of the entropy function, and this would give a real interest to the weak LDP. Unfortunately, even in our framework, the *goodness* of the rate function does not imply a full LDP with this good rate function. It is possible that $\inf_A H(A\|P) > 0$, in which case a full LDP cannot hold, H being nevertheless a good rate function.
Then, some directions may be explored.

1 When is the rate function H a *good* rate function?

2 What are the precise relationships between goodness, exponential tightness, and fullness of the LDP?

3 To which extent is a full LDP necessary?

Partial answers to the first question do exist, e.g. uniformity conditions which can be found in [14, 23, 32, 41]. A general *characterization* of goodness seems difficult to obtain, but perhaps some more specialized results might include bounded jump networks. In fact, for almost all open networks, there is no full LDP, due to the strong homogeneity.

The second question is open. It is known that the exponential tightness implies a full LDP holds with good rate function H (see [14] p. 8). One can conjecture the following: if H is good *and* $\inf_A H(A\|P) = 0$, then the family $\{P^{(n)}, n \geq 0\}$ is exponentially tight. Section 3.2 gives a partial answer to the third point.

3.4 Further extensions

The LDP proved in the preceding sections extends directly to k-tuples empirical measures, to process level, or by considering continuous time or larger state spaces (e.g. Polish spaces).
On the other hand, the entropy function contains some surprises. Actually, when the chain is transient, the global infimum of H can be strictly positive. This infimum is related to the convergence parameter of P

$$\inf H(.\|P) = -\log R(P),$$

where

$$R(P) = \inf_{\lambda}\{\lambda : Ph = \lambda h, \text{ with } h \geq 0\},$$

and this leads naturally to harmonic analysis.

On the other hand, extension of the LDP upper bound to closed sets faces a serious problem. It appears to be closely related to boundary theory, and more precisely to space-time Martin boundary, that is again to harmonic functions. In order to keep the LDP formulation in terms of minima, the entropy function should be extended to the (minimal) Martin boundary, by means of an expression of the form

$$H(h) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{t_n} \log h(x_n, t_n), \tag{3.11}$$

where h is a minimal harmonic function and $(x_n, t_n) \stackrel{n\to\infty}{\longrightarrow} h$ for the Martin compactification. The difficulty is that the limit does not exist, although equation (3.11) makes sense in some concrete situations.

3.5 Identification of the rate function

There are several ways of representing the rate function in Sanov's theorem for the one-dimensional empirical measures

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Actually, the formulations depend on the spaces of $E \mapsto \mathbb{R}$ functions usually considered: $C_b(E)$ (bounded continuous functions), $L_b(E)$ (bounded Lipschitz functions), and also $U(E)$, the space of continuous bounded functions h with $\inf h > 0$. Note that in discrete spaces, all these functions are continuous. The rate functions are given in terms of transforms of the quantities

$$
\begin{aligned}
K_f(u)(x) &\stackrel{\text{def}}{=} \int P(x,\, \mathrm{d}y) e^{f(y)} u(y), \\
T_f(u)(x) &\stackrel{\text{def}}{=} e^{f(x)} \int P(x,\, \mathrm{d}y) u(y), \\
\Lambda(f) &\stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E} \left[\exp \sum_{i=1}^{n} f(X_i) \right],
\end{aligned}
$$

which are related by the relation

$$
\mathbb{E}_x \left[\exp\Bigl(\sum_{i=1}^{n} f(X_i)\Bigr) u(X_n) \right] = e^{f(x)} K_f^{(n-1)}(u)(x) = T_f^{(n-1)}(ue^f)(x). \tag{3.12}
$$

An almost exhaustive list of the main rate functions is given below.

$$
\begin{cases}
I_0(\mu) = \inf\limits_{A^{(1)}=\mu} H(A\|P), \\
I_1(\mu) = \sup\limits_{h\in U(E)} \sum\limits_{i\in E} \mu_i \log \left(\dfrac{h(i)}{Ph(i)} \right), \\
I_2(\mu) = \sup\limits_{f\in C_b(E)} \left[\langle f, \mu \rangle - \Lambda(f) \right], \\
I_3(\mu) = \sup\limits_{f\in L_b(E)} \left[\langle f, \mu \rangle + \log R(K_f) \right], \\
I_4(\mu) = \sup\limits_{f\in L_b(E)} \left[\langle f, \mu \rangle - \log r(T_f) \right].
\end{cases}
$$

where r denotes the spectral radius and R the convergence parameter.

- I_0 is the rate function obtained in Theorem 3.4, and is known to be the same both for the upper and lower bounds.
- I_1 is convenient for computational purposes. It often gives the final form of the result, although it does not appear at first sight in the proofs.
- I_2 is a classical mean of identifying the good rate function whenever the uniformity condition equation (1.1) holds (see [16, 14]). Unfortunately, when working on $\inf_{x\in E} \mathbb{P}_x [L_n \in .]$, the subadditivity method cannot be transposed into a general setting.

- As mentioned earlier, the regeneration argument used in [18, 12]) for the quantity $\mathbb{P}[L_n \in B(\mu,\varepsilon), X_n \in C]$ yields I_3. Getting rid of C is easily done for the lower bound, but is a stumbling block for the upper bound. I_3 also appears in [17] to determine the rate function when there is a full LDP.

- In [10], the inequality $\Lambda(f) \leq r(K_f)$ allows to derive I_4 for the upper bound, but under very stringent conditions regarding the uniformity (see (1.1)).

In the very interesting paper [18], all the above expressions, but for I_0, are shown to be equal, assuming only (roughly speaking) P is irreducible. However, since the goal of the authors is not really to prove a weak LDP, they obtain an upper bound of the form

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[L_n \in K, X_n \in C] \leq - \inf_{\mu \in K} I_3(\mu),$$

for all closed convex sets K, C being typically a non-empty compact set. Moreover, the above upper bound is only proved for exponentially tight sequences $\mathbb{P}[L_n \in .]$. Theorem 1.1 shows that I_0 is in fact valid for both bounds.

In our opinion, an interesting part of our study is to prove the rate function to be the best possible. As a matter of fact, the rate function is unique when a LDP holds[7]. If one aims at getting only one bound, the question is open to decide whether the rate function is optimal or not.

Remark : Clearly, I_0 is equal to the other quantities. For finite support measures μ, an optimization problem, with the $h(i)$ as Lagrange's multipliers, yields $I_0(\mu) = I_1(\mu)$. Now, by the continuity property in Proposition B.3, it is not difficult to check that I_0 is the least lower semi-continuous function such that $I(\mu) = I_0(\mu)$, for all such μ.

Hence $I_0 \leq I_1$, which can also be deduced from regeneration arguments. The reverse inequality is less immediate. Let us denote by L_n^1 [resp. L_n^2] the empirical [resp. pair empirical] measure.

Starting from $\mathbb{P}\left[L_n^2 \in B(A,\varepsilon), X_n = x\right]$, for any fixed x with $a_x > 0$, one carries out the lower bound proof by extending the result from finite support irreducible to any A. The rate function is again $H(A\|P)$, since, from the ergodicity $\mathbb{P}^*\left[L_n^2 \in B(A,\varepsilon), X_n = x\right] \overset{n\to\infty}{\longrightarrow} a_x$.

Fixing $\mu \in M_1(E)$, we have, for any $A^{(1)} = \mu$,

$$L_n^2 \in B(A,\varepsilon) \implies L_n^1 \in B(\mu,\varepsilon).$$

[7]The *usual* argument to prove uniqueness does not apply here, since the LDP is weak and $M_s(E^2)$ is not locally compact. The uniqueness is derived by inequality (3.1)

Hence,

$$\mathbb{P}\left[L_n^2 \in B(A,\varepsilon), X_n = x\right] \leq \mathbb{P}\left[L_n^1 \in B(\mu,\varepsilon), X_n = x\right],$$

which gives

$$H(A\|P) \geq I_1(\mu), \quad \forall A \,:\, A^{(1)} = \mu.$$

Thus $\inf_{A^{(1)}=\mu} H(A\|P) = I_0(\mu) \geq I_1(\mu)$, which is the expected result.

Note that the machinery has to be re-run, in order to show that the rate function I_1 is not too large, which was a priori not obvious for measures with infinite support.

Appendix

A The reducible case

For the sake of completeness, we give in this section a LDP for reducible kernels. No concept is really new, nor is the result surprising, but there is a dense notational thicket hiding the simple idea stated now: instead of evolving ad æternam with the transitions $A = (a_{ij})$, under the twisted probability $\mathbb{P}^*$ the chain behaves for a long time like $A(1)$, then like $A(2)$, and so on, where the $A(k)$ correspond to different classes of irreducibility.
In fact, in this case, the bounds are obtained by a change of measure which is *not time-homogeneous*, and the subset $\mathcal{M}$ is not convex nor is the rate function, thus rendering the convex analysis ineffectual.

Theorem A.1 (reducible case) *Let X be a Markov chain with kernel P, and let $\mathcal{M}$ be the set of serial measures relatively to P. The pair empirical measure of X satisfies a weak large deviation principle, with the rate function*

$$H_{|\mathcal{M}}(A\|P) \stackrel{\text{def}}{=} \begin{cases} H(A\|P), & \textit{for } A \in \mathcal{M}; \\ \infty, & \textit{otherwise.} \end{cases} \tag{A.1}$$

A.1 Irreducible classes and associated objects

Throughout this section, P will be a reducible transition matrix. When $P^{(n)}(x,y) > 0$, there exits a path of length n, of positive probability, connecting the points x and y, and we shall write $x \rightsquigarrow y$. Two points belong to the same class of irreducibility if, and only if, $x \rightsquigarrow y$ and $y \rightsquigarrow x$.
The state space is partitioned into classes of irreducibility $E = \bigcup_{i\in\mathcal{I}} E_i$, the graph of P on E_i being denoted by G_i. Then there is the following classical partial ordering $\succ$ between the classes: $E_j \succ E_i$ if there exist $x \in E_i$ and $y \in E_j$, such that $x \rightsquigarrow y$. Thus $x \rightsquigarrow y$ for all $x \in E_i$ and $y \in E_j$. Eventually, the points reachable from $x \in E_i$ are all the sets E_j with $E_j \succ E_i$. The forthcoming definition is a formal description of the “suitable” measures.

Definition A.1 (serial measures) *A* serial *measure A relatively to P is a balanced measure*

$$A = \sum u_i A(i), \ \textit{with } A(i) \in M_s(G_i) \ \textit{ and } E_{i+1} \succ E_i, \ \forall i.$$

In addition it is required that $\nu\big(\{x \in E : E_1 \succ x\}\big) > 0$*, where* ν *is the initial distribution of the chain. The set of serial measures will be denoted by* $\mathcal{M}$*.*

Lemma A.2 *Serial measures are characterized by the equivalence*

$$A \in \mathcal{M} \iff \left\{ \begin{array}{l} \exists A(m) \stackrel{m\to\infty}{\longrightarrow} A, \\ \mathbb{P}\left[L_{n(m)} = A(m)\right] > 0. \end{array} \right. \tag{A.2}$$

Proof : The sufficient condition $\Longleftarrow$
Let i, j be such that $a_{ij} > 0$. This implies $a_{ij}(m) > 0$ for large m , so that $i \rightsquigarrow j$, and, by the balance assumption, $j \rightsquigarrow i$ also. Therefore A can be split into irreducible classes, and one can write

$$A = \sum u_i A(i), \quad \text{with } A(i) \in M_s(G_i).$$

Take now $j_0 \in E_j$ and $k_0 \in E_k$, with $a_{j_0}a_{k_0} > 0$, so that $u_j u_k > 0$. Consequently, for m sufficiently large, we have $a_{j_0}(m)a_{k_0}(m) > 0$, and a path crosses j_0 and k_0 with a positive probability. So, either $E_j \succ E_k$ or $E_k \succ E_j$, which means that $\succ$ is a total order relation on the sets E_i for which $u_i > 0$. Consequently, A is serial.

The Necessary condition $\Longrightarrow$
Assume first A has a finite support and construct the sequence $A(m)$. The decomposition of A as a serial measure is finite and so is its decomposition into cycles (see[8] Proposition C.2). Hence $A = \sum_{i=1}^{p} u_i L(C_i)$, and cycles can be ordered so that $C_{i+1} \succ C_i$. There exist paths $T_{i,i+1} = (x_{n_i}, \dots, x_{n_{i+1}})$, with positive probability, such that $x_{n_i} \in C_i$, for $i \geq 1$, and $\nu(x_{n_0}) > 0$. Then, taking integer approximations $\alpha_i(m)$ of mu_i, modified by the length of C_i, the path

$$T(m) = T_{01} C_1^{\alpha_1(m)} T_{12} C_2^{\alpha_2(m)} T_{23} \dots C_p^{\alpha_p(m)}$$

has also a positive probability. Finally $A(m) \stackrel{\text{def}}{=} L(T(m))$ converges to A when m grows and, by construction, $L_{n(m)} = A(m)$.
In the case A has an infinite support, there exists a sequence of serial $A(n)$ with finite support converging to A. For each $A(n)$, the above paragraph shows there exists a sequence $A(n, m)$ converging to $A(n)$, with positive probability.
Finally, letting $A'(n) \stackrel{\text{def}}{=} A(n, m_n)$, with m_n such that

$$\|A'(n) - A(n)\| \leq 1/n,$$

one sees that the sequence $A'(n)$ also converges to A, and the proof is concluded. ■

[8]Note that Proposition C.2 does not postulate irreducibility.

A.2 Lower bound

Theorem *For any open set $O \subset M_s(E^2)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in O\right] \geq - \inf_{A\in O\cap\mathcal{M}} H(A\|P). \tag{A.3}$$

Proof : The argument is like in the irreducible case, but classes have to be separated. Let $A \in O \cap \mathcal{M}$ be a measure with finite support, so that the decomposition $A = \sum_{i=1}^{p} u_i A(i)$ is finite. Define

$$h_k(i,j) \stackrel{\text{def}}{=} \begin{cases} \log \frac{A_{ij}(k)}{P_{ij}}, & \text{if } i,j \in \operatorname{Supp}(A(k)), \\ 0 & \text{otherwise.} \end{cases} \tag{A.4}$$

Fix a large integer N, take $n \leq N$, and let

$$t_0 = 0 \quad \text{and} \quad t_i = \lfloor N(u_1 + \cdots + u_i) \rfloor \wedge n.$$

The martingale M_n of equation (2.2) is replaced by a sequence of martingales

$$M_n^{(N)} \stackrel{\text{def}}{=} \exp\left\{ \sum_{k=1}^{p} \sum_{i=t_{k-1}+1}^{t_k} h_k(X_i, X_{i+1}) \right\}. \tag{A.5}$$

By the definition of t_k, the summation is taken over the n first terms at most. It is easily checked that $M_n^{(N)}$ is a martingale, for $n \leq N$, defining a new measure $\mathbb{P}^*$. Under $\mathbb{P}^*$, X is a Markov chain, which is no longer time homogeneous: it behaves like $A(i)$ on $\operatorname{Supp}(A(i))$ and like P outside, for a duration $[t_{i-1}+1, t_i]$, for all $i \geq 1$. For each k, we introduce the partial empirical measures

$$L_n^k \stackrel{\text{def}}{=} \frac{1}{t_k} \left(\sum_{i=t_{k-1}+1}^{t_k - 1} \delta_{X_i X_{i+1}} + \delta_{X_{t_k} X_{t_{k-1}+1}} \right) \in M_s(G_k). \tag{A.6}$$

Choose an open neighborhood $\mathcal{V}_A \subset O$ of A. For n sufficiently large, the ghost transitions are small enough, since their total mass being less than k/n. Hence, there exist open neighborhoods $\mathcal{V}_A(k)$ of $A(k)$, with

$$L_n^k \in \mathcal{V}_A(k),\ \forall k \implies L_n \in \mathcal{V}_A. \tag{A.7}$$

Combining (2.3) and (A.7), yields, when $N = n$,

$$
\begin{aligned}
\frac{1}{n}\log \mathbb{P}\left[L_n \in \mathcal{V}_A\right] \geq \frac{1}{n}\log \mathbb{E}^*\left[\prod_{k=1}^{p} \mathbb{1}_{\{L_n^k \in \mathcal{V}_A(k)\}}\left(M_n^{(n)}\right)^{-1}\right] & \\
\geq \sum_{k=1}^{p} -\frac{t_k}{n} \sup_{A' \in \mathcal{V}_A(k) \cap M_s(G'_k)} \sum_{(i,j)\in G'_k} a'_{ij} h_k(i,j) & \qquad \text{(A.8)}\\
+ \frac{1}{n}\log \mathbb{E}^*\left[\prod_{k=1}^{p} \mathbb{1}_{\{L_n^k \in \mathcal{V}_A(k)\}} \exp\big(h_k(X_{t_k}, X_{t_{k-1}+1})\big)\right], & \qquad \text{(A.9)}
\end{aligned}
$$

where G'_k reads both for the graph and the support of $A(k)$. By continuity, the term in (A.8) gives $u_k H(A(k)\|P)$, and hence the sum tends to $-H(A\|P)$, for $n \to \infty$ and $\mathcal{V}_A$ small enough.
The expectation appearing in (A.9) is not that complicated. The k-th functional indeed depends only on X_i, for $i \in [t_{k-1}+1, t_k]$, and X behaves like $A(k)$ on this time interval. The chain has been split into p pieces, to each of which the result of the irreducible case can be applied. But the chain must hit once G'_k, and only after this moment it behaves like $A(k)$. Remembering that $\mathbb{P}^*$ depends on n, it is not obvious whether the probability of ever reaching G'_k under $\mathbb{P}^*$ remains uniformly bounded on the sets of interest.

One will show by induction that there exist n_k and $b_k > 0$, such that

$$
\mathbb{P}^*\left[X_{t_{l-1}+n_l} \in G'_l,\ \forall l \leq k\right] \geq b_k > 0, \qquad \text{(A.10)}
$$

independently of n. Since $\nu(\{x \in E:\ E_1 \succ x\}) > 0$, the assertion is true for $k = 1$. Assume it is true for k. Then

$$
\mathbb{P}^*\left[X_{t_{l-1}+n_l} \in G'_l,\ \forall l \leq k \text{ and } X_{t_k} \in G'_k\right] \geq b_k, \qquad \text{(A.11)}
$$

as, under $\mathbb{P}^*$, G'_k is absorbing during the time interval $[t_k+1, t_{k+1}]$. Choose n_{k+1} such that

$$
\inf_{x \in G'_k} \mathbb{P}^*\left[X_{t_k+n_{k+1}} \in G'_{k+1} | X_{t_k} = x\right] = c_k > 0. \qquad \text{(A.12)}
$$

This can be done since $E_{k+1} \succ E_k$ and G'_k is finite. Thus, applying the Markov property to equations (A.11)–(A.12),

$$
\mathbb{P}^*\left[X_{t_{l-1}+n_l} \in G'_l,\ \forall l \leq k+1\right] \geq b_k c_k = b_{k+1} > 0, \qquad \text{(A.13)}
$$

and (A.10) holds for $k+1$.
Now, one will use the ergodic theorem on the interval $[t_{k-1}+n_k, t_k]$. Since $n_1+\cdots+n_p$ is constant, independently of n, it is easy to modify the partial empirical measures (see (A.6)) to ensure finally

$$\liminf_{n\to\infty} \mathbb{E}^* \left[\prod_{k=1}^{p} \mathbb{1}_{\{L_n^k \in \mathcal{V}_A(k)\}} e^{h_k(X_{t_k}, X_{t_{k-1}+n_k})} \right] \geq b_p \prod_{k=1}^{p} \inf_{(i,j)\in G'_k} e^{h_k(i,j)} > 0,$$

because G'_k is finite, so that the infima are strictly positive. Therefore the (modified) sum of line (A.9) tends to 0 when $n \to \infty$.

For measures with infinite supports, approximations and continuity arguments [Proposition B.3 still holds for reducible kernels] yield the result along the same way as in the irreducible case. ■

> **Remark :** Note that the proof would have been more difficult in taking directly infinite support measures, for uniformity reasons.
>
> The above construction can be related to the control and cost structure described in [20], but we do not need here to deal with all control policies, since it is possible to choose directly the optimal one. However the aims are somewhat different and the change of measure allows to obtain the rate function in a closed form, instead of an expression as a "limit of limit".

A.3 Upper bound

Theorem *For any compact set $K \subset M_s(E^2)$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A\in K\cap\mathcal{M}} H(A\|P). \tag{A.14}$$

Proof : It is much simpler. When A is not serial, by Lemma A.2,

$$A \notin \mathcal{M} \implies \exists \mathcal{V}_A,\ \exists n_0 : \forall n \geq n_0, \quad \mathbb{P}\left[L_n \in \mathcal{V}_A\right] = 0,$$

where $\mathcal{V}_A$ is an open neighborhood of A. Then $\mathbb{P}\left[L_n \in K \setminus \mathcal{V}_A\right]$ equals $\mathbb{P}\left[L_n \in K\right]$, for n large enough, and the upper bound becomes

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A'\in K\setminus\mathcal{V}_A} H(A'\|P).$$

This can be achieved by means of finite number of neighborhoods. Using the lower semi-continuity of H and the closure of $K \cap \mathcal{M}$, we get

$$\inf \left\{ \inf_{A' \in K \setminus \cup_{i=1}^{p} \mathcal{V}_{A_i}} H(A' \| P) \ , \ A_1, \ldots, A_p \notin \mathcal{M} \right\} = \inf_{A' \in K \cap \mathcal{M}} H(A' \| P),$$

which gives the upper bound (A.14). ■

B Properties of the relative entropy function

The entropy function lies in the heart of the LDP, and we need to analyze some of its properties to derive the large deviation assertion. In this section, P is again a Markov kernel on the countable state space E.

B.1 Entropy and its information theoretic interpretation

Definition B.1 (balanced measures) *The set $M_s(E^2)$ of balanced measures is the set of measures on E^2, with both identical 1–dimensional projections*

$$A \in M_s(E^2) \iff A(E, .) = A(., E) \in M_1(E). \tag{B.1}$$

Let G be a graph on E. $M_s(G) \subset M_s(E^2)$ is the set of balanced measures with support included in G.

For $A \in M_s(E^2)$, let us put

- $a_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\})$ the 2–dimensional law;
- $a_i \stackrel{\text{def}}{=} A(\{i\} \times E)$, the unique 1–dimensional projection, also denoted by $A^{(1)}$;
- $A_{ij} \stackrel{\text{def}}{=} A(E \times \{j\} | \{i\} \times E)$, the conditional law, and hence $a_{ij} = a_i A_{ij}$.

Definition B.2 (relative entropy) *Let $A \in M_s(E^2)$. The relative entropy of A with respect to P is defined by*

$$H(A \| P) \stackrel{\text{def}}{=} \sum_{i,j \in E} a_{ij} \log \left(\frac{a_{ij}}{a_i P_{ij}} \right), \tag{B.2}$$

with the usual convention

$$0 \log(0/0) = 0, \quad 1/0 = +\infty, \quad 0 \log 0 = 0.$$

This definition extends in a natural way to finite positive measures A and substochastic matrices P. $H_{E'}$ is the restriction of the entropy H to E', i.e. taking the indices $i, j \in E'$ in equation (B.2).

All results, and the next as well, do have an information theory significance. The reader is referred to Rényi [37], which provides an excellent introduction containing the link between information theory and probability. For a more complete study, see [1], and Hamming [25].
Recalling that the *information* brought by a Bernoulli trial Y is $-\log \mathbb{P}[Y]$, one sees $-n^{-1}\log \mathbb{P}[L_n \in F]$ as the mean information brought by the knowledge of L_n being in F. By the way, we think the expression "relative entropy" not to be as illuminating as *information gain*. Indeed the relative entropy is the *mean information gain*:

$$H(A\|P) = \sum_{i\in E} a_i \sum_{j\in E} A_{ij} \log\left(\frac{A_{ij}}{P_{ij}}\right) = \mathbb{E}^{A^{(1)}}\left[I(A_{X_1}\|P_{X_1})\right], \tag{B.3}$$

where $I(A_{X_1}\|P_{X_1}) \stackrel{\text{def}}{=} \sum_j A_{ij}\log(A_{ij}/P_{ij})$ is the information gain[9] for X_2, conditioning on the event $\{X_1 = i\}$. Thus $H(A\|P)$ is the mean information gain for each sample, under the dynamics A.

Lemma B.1 *The relative entropy $H(A\|P)$ is positive, convex and lower semi-continuous with respect to A. It is null if, and only if, one can write $A = P$ as a Markov kernel, i.e. $A_{ij} = P_{ij}$, $\forall i, j \in E$.*

Proof : Positivity, lower semi-continuity and convexity are immediate from equation (B.3).
As for the convexity, let $A, A' \in M_s(E^2)$, $x, x' \geq 0$ with $x + x' = 1$. Then $B = xA + x'A'$ is also balanced with

$$b_i = xa_i + x'a'_i, \quad B_{ij} = (xa_{ij} + x'a'_{ij})/(xa_i + x'a'_i),$$

so that

$$\begin{aligned} xH(A\|P) + x'H(A'\|P) &= \sum_{i\in E} xa_i I(A_i\|P_i) + x'a'_i I(A'_i\|P_i) \\ &\geq \sum_{i\in E} I(xa_iA_i + x'a'_iA'_i\|(xa_i + x'a'_i)P_i) \\ &= \sum_{i\in E} b_i I(B_i\|P_i) = H(B\|P), \end{aligned}$$

[9] Here, one writes I, but this is actually the relative entropy of measures in $M_1(E)$.

by the convexity and the 1–homogeneity of I. ■

We will state now an important *negative* property. Let

$$\Psi(K) \stackrel{\text{def}}{=} \{A \in M_s(E^2) : H(A\|P) \le K\}$$

denote the level set of the entropy function.

The closed level set $\Psi(K)$ is not necessarily compact.

Proof : Consider the irreducible ergodic kernel P defined on $\mathbb{N}$ by

$$P_{01} = 1, \qquad \text{and} \qquad P_{ij} = \begin{cases} 3/8 & \text{if} \quad j = i-1, \\ 1/2 & \text{if} \quad j = i, \\ 1/8 & \text{if} \quad j = i+1, \end{cases} \qquad \text{for } i \ge 1.$$

Then, for each $q > 0$,

$$H(\delta_{qq}\|P) = -\log P_{qq} = \log 2.$$

Hence, the set $\{\delta_{qq} : q \ge 1\} \subset \Psi(\log 2)$ is not tight and $\Psi(\log 2)$ is not compact. ■

> **Remark :** Consequently, a full LDP will, in general, not hold, since it would require the entropy function to have compact level sets. In this paper, we prove that there is a *weak* LDP for *all* Markov chains. Obviously, a full LDP will hold only for some restricted classes of chains.

B.2 Continuity properties of the entropy function

The proof of the lower bound of the LDP needs continuity properties of the entropy function. According to (Definition C.1), the graph G will be assumed, in this section, to correspond to an irreducible transition matrix P. Note that $H(A\|P)$ is finite only if $A \in M_s(G)$.

Proposition B.2 (continuity) *Let G' be a finite subgraph of G. Then $H(.\|P)$ is finite and continuous on $M_s(G')$.*

Proof : The entropy function is a finite sum, for $(i,j) \in G'$, of finite continuous functions $A \mapsto a_{ij} \log(a_{ij}/a_i P_{ij})$, since $P_{ij} > 0$, $\forall i, j$. ■

Proposition B.3 (exact l.s.c.) *Let $A \in M_s(G)$. There exists a sequence of measures $A(n) \in M_s(G)$, with finite support, converging to A, such that their entropy converges to $H(A\|P)$.*

$$\lim_{n\to\infty} A(n) = A \quad \textit{and} \quad \lim_{n\to\infty} H(A(n)\|P) = H(A\|P).$$

Proof : Assume $H(A\|P) < \infty$. By Proposition C.2, there exists a decomposition into cycles $A = \sum_{i=1}^{\infty} u_i L(C_i)$, from which one can define the sequence

$$A(n) \stackrel{\text{def}}{=} \sum_{i=1}^{n} u_i L(C_i) + \sum_{i=n+1}^{\infty} u_i L(C_1) \ \in M_s(G) \ \stackrel{n\to\infty}{\longrightarrow} \ A. \tag{B.4}$$

In addition, let $K \stackrel{\text{def}}{=} \mathrm{Supp}(C_1)$. Since K is finite, there exists q_0 such that

$$K \subset \{1, \dots, q_0\}^2.$$

Introduce the measure $\pi_{ij} \stackrel{\text{def}}{=} a_i P_{ij}$ and the ratios

$$\begin{cases} r_{ij} \stackrel{\text{def}}{=} a_{ij}/\pi_{ij}, \\ r_{ij}(n) \stackrel{\text{def}}{=} a_{ij}(n)/\pi_{ij}, \\ r_i(n) \stackrel{\text{def}}{=} a_i(n)/a_i. \end{cases}$$

By construction,

$$-e^{-1} \le r_{ij}(n) \log r_{ij}(n) \le \max\{0, r_{ij} \log r_{ij}\}, \quad \forall (i,j) \notin K, \tag{B.5}$$

$$-e^{-1} \le r_i(n) \log r_i(n) \le 0, \qquad \forall i > q_0. \tag{B.6}$$

These inequalities are derived from special forms of the decomposition equation (B.4). Indeed, writing for the sake of shortness $f(x) = x \log x$,

$$\begin{aligned} H(A(n)\|P) \quad = \quad & \sum_{(i,j)\in K} \pi_{ij} f\big(r_{ij}(n)\big) - \sum_{i \le q_0} a_i f\big(r_i(n)\big) \\ & + \sum_{(i,j)\in K^c} \pi_{ij} f\big(r_{ij}(n)\big) - \sum_{i > q_0} a_i f\big(r_i(n)\big) \end{aligned} \tag{B.7}$$

The first two sums in equation (B.7) are finite and converge, because f is continuous and there is finite number of terms. Similarly, using the bounds in (B.5)–(B.6), one sees the last two sums in (B.7) are also finite. Thus, Lebesgue's theorem implies that

$$H(A(n)\|P) \stackrel{n\to\infty}{\longrightarrow} H(A\|P).$$

When $H(A\|P) = \infty$, then a finite number of terms are sufficient to prove that $H(A(n)\|P)$ is larger than any fixed number, so that the limit is still true. The proof of Theorem B.3 is terminated. ■

C Cycles and decomposition of balanced measures

The main goal of this section is to provide in Proposition C.2 an effective decomposition of balanced measure, in order to analyze the continuity of the entropy function. To this end, convenient tools are graph theory [3], convex sets properties [38, 2, 28] and the bible [6].
A graph will be described as a subset G of $E \times E = E^2$, where E is countable. This appendix is independent of Markov chains, although all graphs considered hereafter are indeed associated with Markov transition matrices.

Definition C.1 *The graph G of a Markov transition matrix P is the set of all possible transitions, i.e.* $(i,j) \in G \iff P_{ij} > 0$.

Clearly, G is strongly connected if, and only if, P is irreducible. A cycle is described as an n-tuple $(x_1, \ldots, x_n)$, with all $(x_i, x_{i+1}) \in G$. It is minimal if, and only if, it passes through each vertex at most once, so that the number of minimal cycles is at most countable. These cycles are denoted by $(C_i)_{i\in\mathbb{N}}$. They will play an important role, essentially because any cycle has at least one minimal cycle. The empirical measure $L(C)$ of a cycle C belongs to $M_s(G)$. The following proposition introduces some approximations that are needed for the method of change of measure.

Proposition C.1 *Let G be a strongly connected graph, $A \in M_s(G)$ with a finite support. There exists a finite subgraph $G' \subset G$ and balanced measures $A(n)$ converging to A, such that* $\operatorname{Supp} A(n)$ *is strongly connected and belongs to G'.*

Proof : Let $\{x_1, \ldots, x_n\}$ be the support of A. Since G is strongly connected, there exists a G–path $T_{i,j}$ from x_i to x_j, hence $C = T_{1,2} \ldots T_{n,1}$ is a cycle connecting all points of the support of A.
Taking $G' = \operatorname{Supp}(C) \cup \operatorname{Supp}(A)$ and $A(n) = n^{-1}L(C) + (1-n^{-1})A$, the proposition is proved. ■

We also need finite approximations. One can easily guess that the $L(C_i)$ generate $M_s(G)$, but, when G is countable, there is also a countable number of minimal cycles, hence $M_s(G)$ is neither finitely generated nor compact, so that a Krein–Millman like theorem must be proved for our special situation and this is the content of the next Proposition C.2.
Convex sets will be considered in a topological vector space denoted by $M_s^{ev}(G)$, the set of *finite signed* balanced measures on G. Endowed with the L_1-norm

$$\|A\| = \sum_{ij} |a_{ij}| < \infty,$$

$M_s^{ev}(G)$ is even a Banach space. The cone $M_s^+(G) \subset M_s^{ev}(G)$ of *positive* balanced measures is the right object to study, since the extreme points of $M_s(G)$ correspond to extreme generatrices of the cone, but we will strive to focus rather on simplicity than on full generality.

Proposition C.2 (balanced measure decomposition) *The set of balanced measures having their support in G is the closed convex hull generated by the empirical measures of the minimal cycles of G, that is*

$$M_s(G) = \overline{\text{co}}\left(\left\{L(C_i)\right\}_{i\in\mathbb{N}}\right). \tag{C.1}$$

Proof : Since $M_s(G)$ is closed, $\overline{\text{co}}(\{C_i\}_{i\in\mathbb{N}}) \subset M_s(G)$, because the closed convex hull is the closure of the convex hull.

Conversely, take $A \in M_s(G)$, and consider $A - uL(C_1)$. This is a balanced measure, positive for $u = 0$. Introduce the sequences

$$\begin{cases} u_{i+1} = \max\left\{u : B_i - uL(C_i) \in M_s^+(G)\right\}, \\ B_{i+1} = B_i - u_i L(C_i), \ \forall i \geq 1, \quad \text{with } B_1 = A. \end{cases} \tag{C.2}$$

This yields a decomposition with non-negative coefficients $u_1, u_2, \ldots$ and

$$B = A - \sum_i u_i L(C_i)$$

is positive, since $M_s^+(G)$ is closed. Suppose now $B \neq 0$. Then there exists x_1, such that $b_{x_1} = \varepsilon > 0$. Since B is positive and balanced,

$$b_x = \sum_j b_{xj} = \sum_i b_{ix}, \ \forall x,$$

so that one can define by induction the sets

$$\mathcal{S}_1 \stackrel{\text{def}}{=} \{x_1\}, \quad \mathcal{S}_{n+1} \stackrel{\text{def}}{=} \{x_{n+1} : b_{x_n x_{n+1}} > 0, \quad \forall x_n \in \mathcal{S}_n\} \neq \emptyset,$$

where

$$\sum_{i \in \mathcal{S}_n} \sum_{j \in \mathcal{S}_{n+1}} b_{ij} \geq \varepsilon. \tag{C.3}$$

If a cycle could be extracted from $\mathcal{S}_1, \mathcal{S}_2, \dots$, then B would possess a minimal cycle C_l, and u_l would not satisfy equation (C.2). On the other hand, if all $\mathcal{S}_n$ would be disjoint, then $\|B\|$ would as large as the quantity obtained by summing equation (C.3) over all $n \geq 1$, which is infinite. Both conclusions are contradictory. Consequently,

$$B = 0, \quad A = \sum_i u_i L(C_i),$$

which implies as well that $M_s(G) \subset \overline{\text{co}}(\{L(C_1), \dots, L(C_p)\})$.
The proof of Proposition C.2 is terminated. ■

D Notation

The following notation is used throughout the paper.

- E is a countable state space isomorphic to $\{1, \dots, |E|\}$.
- $\{X_k, k \geq 1\}$ is a Markov chain on E, with transition matrix $P = (P_{ij})$.
- $M_s(E^2)$, introduced in Definition B.1, is the set of balanced measures on $E \times E = E^2$. For any $A \in M_s(E^2)$,
  - $a_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\})$ is the 2–dimensional law;
  - $A^{(1)}$ is the 1–dimensional projection, with
  $$a_i \stackrel{\text{def}}{=} A^{(1)}(\{i\}) = A(\{i\} \times E) = A(E \times \{i\});$$
  - A_{ij}, is the conditional law, so that $a_{ij} = a_i A_{ij}$;
  - $\text{Supp}(A) \subset E^2$ is the support of A.
- G is a graph, $M_s(G)$ is the set of balanced measures on this graph.
- $L_n(\omega)$ is the pair empirical measure (Definition 1.1).
- $H(A\|P)$ is the relative entropy of A with respect to P (Definition B.2).

References

[1] Ash, R. *Information theory.* Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1965. Interscience Tracts in Pure and Applied Mathematics, No. 19.

[2] Barbu, V., and Precupanu, T. *Convexity and optimization in Banach spaces*, second ed., vol. 10 of *Mathematics and its Applications (East European Series).* D. Reidel Publishing Co., Dordrecht, 1986.

[3] Berge, C. *Graphs*, vol. 6 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, 1985. Second revised edition of part 1 of the 1973 English version.

[4] Bolthausen, E. Markov process large deviations in τ-topology. *Stochastic Process. Appl. 25*, 1 (1987), 95–108.

[5] Bolthausen, E., and Schmock, U. On the maximum entropy principle for uniformly ergodic Markov chains. *Stochastic Process. Appl. 33*, 1 (1989), 1–27.

[6] Bourbaki, N. *Éléments de mathématique. Fasc. XV. Livre V: Espaces vectoriels topologiques. Chapitre I: Espaces vectoriels topologiques sur un corps valué. Chapitre II: Ensembles convexes et espaces localement convexes.* Hermann, Paris, 1966. Actualités Scientifiques et Industrielles, No. 1189. Deuxième édition revue et corrigée.

[7] Cattiaux, P., and Léonard, C. Minimization of the Kullback information for some Markov processes. In *Séminaire de Probabilités, XXX*, vol. 1626 of *Lecture Notes in Math.* Springer, Berlin, 1996, pp. 288–311.

[8] Csiszár, I., Cover, T. M., and Choi, B. S. Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory 33*, 6 (1987), 788–801.

[9] Csiszár, I., and Körner, J. *Coding theorems for discrete memoryless systems.* Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

[10] de Acosta, A. Upper bounds for large deviations of dependent random vectors. *Z. Wahrsch. Verw. Gebiete 69*, 4 (1985), 551–565.

[11] DE ACOSTA, A. Large deviations for empirical measures of Markov chains. *J. Theoret. Probab. 3*, 3 (1990), 395–431.

[12] DE ACOSTA, A., AND NEY, P. Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab. 26*, 4 (1998), 1660–1682.

[13] DE LA FORTELLE, A., AND FAYOLLE, G. Large deviation principle for a markov chain with a countable state space. In *Rapport de recherche*, vol. 3503. INRIA, 1998, p. 39.

[14] DEMBO, A., AND ZEITOUNI, O. *Large deviations techniques and applications.* Jones and Bartlett Publishers, Boston, MA, 1993.

[15] DEMBO, A., AND ZEITOUNI, O. *Large deviations techniques and applications*, second ed. Springer-Verlag, New York, 1998.

[16] DEUSCHEL, J.-D., AND STROOCK, D. W. *Large deviations*, vol. 137 of *Pure and Applied Mathematics.* Academic Press Inc., Boston, MA, 1989.

[17] DINWOODIE, I. H. Identifying a large deviation rate function. *Ann. Probab. 21*, 1 (1993), 216–231.

[18] DINWOODIE, I. H., AND NEY, P. Occupation measures for Markov chains. *J. Theoret. Probab. 8*, 3 (1995), 679–691.

[19] DONSKER, M. D., AND VARADHAN, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time. I, II, III. *Comm. Pure Appl. Math. 28* (1975, 1976), 1–47; **28** 279–301; **29** 389–461.

[20] DUPUIS, P., AND ELLIS, R. S. The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc. 347*, 8 (1995), 2689–2751.

[21] ELLIS, R. S. Large deviations for a general class of random vectors. *Ann. Probab. 12*, 1 (1984), 1–12.

[22] ELLIS, R. S. *Entropy, large deviations, and statistical mechanics*, vol. 271 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science].* Springer-Verlag, New York, 1985.

[23] ELLIS, R. S. Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probab. 16*, 4 (1988), 1496–1508.

[24] GALLAGER, R. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[25] HAMMING, R. W. *Coding and information theory.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[26] JAIN, N. C. Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab. 18*, 3 (1990), 1071–1098.

[27] KULLBACK, S., AND LEIBLER, R. A. On information and sufficiency. *Ann. Math. Statistics 22* (1951), 79–86.

[28] LAY, S. R. *Convex sets and their applications.* Robert E. Krieger Publishing Co. Inc., Malabar, FL, 1992. Revised reprint of the 1982 original.

[29] LEWIS, J., AND PFISTER, C. Thermodynamic probability theory: some aspects of large deviations. *Uspekhi Mat. Nauk 50*, 2(302) (1995), 47–88.

[30] LINDQVIST, B. On the loss of information incurred by lumping states of a Markov chain. *Scand. J. Statist. 5*, 2 (1978), 92–98.

[31] LINNIK, J. V. An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Teor. Verojatnost. i Primenen 4* (1959), 311–321.

[32] LIPTSER, R. S. Large deviations for empirical measures of Markov processes (discrete time, the noncompact case). *Teor. Veroyatnost. i Primenen. 41*, 1 (1996), 65–88.

[33] MICLO, L. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In *Séminaire de Probabilités, XXXI*, vol. 1655 of *Lecture Notes in Math.* Springer, Berlin, 1997, pp. 136–167.

[34] MILGRAM, M. Barycentric entropy of Markov systems. *Cybernet. Systems 12*, 1-2 (1981), 141–178.

[35] MILLER, H. D. A convexivity property in the theory of random variables defined on a finite Markov chain. *Ann. Math. Statist. 32* (1961), 1260–1270.

[36] O'BRIEN, G. L., AND VERVAAT, W. Capacities, large deviations and loglog laws. In *Stable processes and related topics (Ithaca, NY, 1990).* Birkhäuser Boston, Boston, MA, 1991, pp. 43–83.

[37] Rényi, A. *Calcul des probabilités. Avec un appendice sur la théorie de l'information.* Dunod, Paris, 1966. Traduit de l'allemand par C. Bloch. Collection Universitaire de Mathématiques, No. 21.

[38] Rockafellar, R. T. *Convex analysis.* Princeton University Press, Princeton, N.J., 1970. Princeton Mathematical Series, No. 28.

[39] Sanov, I. N. On the probability of large deviations of random variables. In *Select. Transl. Math. Statist. and Probability, Vol. 1.* Inst. Math. Statist. and Amer. Math. Soc., Providence, R.I., 1961, pp. 213–244.

[40] Shwartz, A., and Weiss, A. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[41] Wu, L. M. Grandes déviations pour les processus de Markov essentiellement irréductibles. I. Temps discret. *C. R. Acad. Sci. Paris Sér. I Math. 312*, 8 (1991), 609–614.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399